Strong convergence of partial match queries in random quadtrees

Nicolas Curien
École Normale Supérieure

Abstract

We prove that the rescaled costs of partial match queries in a random two-dimensional quadtree converge *almost surely* towards a random limit which is identified as the terminal value of a martingale. Our approach shares many similarities with the theory of self-similar fragmentations.

1 Introduction

The quadtree structure is a storage system designed for retrieving multidimensional data. It has first been introduced by Finkel & Bentley [7] and was studied thoroughly in computer science. The goal of this work is to study fine properties of the so-called partial match queries in random (uniform) two-dimensional quadtrees.

Let us briefly recall the model. Consider a Poisson point process Π on $\mathbb{R}_+ \times [0,1]^2$ with intensity $\mathrm{d}t \otimes \mathrm{d}x\mathrm{d}y$. Let $((\tau_i, x_i, y_i), i \geq 1)$ be the atoms of Π ranked in the increasing order of their τ-component. We define a process $(\mathrm{Quad}(t))_{t\geq 0}$ with values in finite coverings of $[0,1]^2$ by closed rectangles with disjoint interiors as follows. We initially start with the unit square $\mathrm{Quad}(0) := [0,1]^2$. At each time an atom (τ_i, x_i, y_i) of the Poisson process Π falls in a rectangle of $\mathrm{Quad}(\tau_i^-)$ it splits this rectangle into four subrectangles according to the horizontal and vertical coordinates of x_i and y_i. Observe that a.s., for every $i \geq 1$, there exists a unique rectangle of $\mathrm{Quad}(\tau_i)$ such that (x_{i+1}, y_{i+1}) is in its interior, hence the process $(\mathrm{Quad}(t))_{t\geq 0}$ is a.s. well defined. In this work, we chose to focus on the continuous time version of the random quadtree but all the results can be transferred to the random quadtree with a fixed number of points by standard depoissonization techniques, see e.g. [3] or [5, Lemma 1].

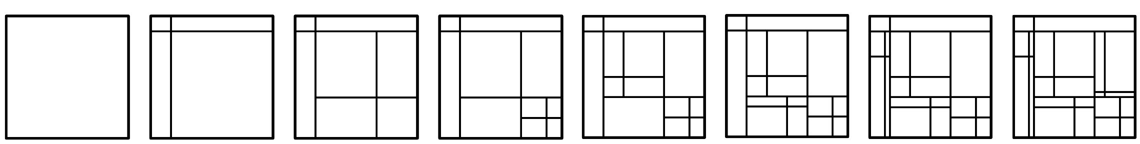

Figure 1: The first 7 splittings of a quadtree.

We shall be interested in the so-called partial match query (see [9, p 523]). Equivalently, for $x \in [0,1]$, we focus on the number $N_t(x)$ of rectangles in the quadtree at time t whose horizontal coordinate intersects x, that is,

$$N_t(x) \quad := \quad \#\big\{R \in \mathrm{Quad}(t) : R \cap [(x,0),(x,1)] \neq \emptyset\big\}.$$

The first study of the partial match has been carried out by Flajolet, Gonnet, Puech and Robson in [8]. They proved that if U is uniformly distributed over $[0,1]$ and independent of Π then $E[N_t(U)]$ is asymptotically equivalent to $\kappa \cdot t^{\beta}$ as t tends to infinity, where

$$\beta \quad := \quad \frac{\sqrt{17}-3}{2},$$

and $\kappa > 0$ is some explicit constant. The asymptotic of the expected value of $N_t(x)$ for a fixed point $x \in [0,1]$ has recently been obtained in [5], it reads

$$t^{-\beta} E[N_t(x)] \quad \xrightarrow[t\to\infty]{} \quad K_0 \cdot h(x), \tag{1}$$

where

$$K_0 := \frac{\Gamma(2\beta+2)\Gamma(\beta+2)}{2\Gamma^3(\beta+1)\Gamma^2(\beta/2+1)} \quad \text{and} \quad h : u \in [0,1] \longmapsto \big(u(1-u)\big)^{\beta/2}.$$

In a very recent breakthrough [4], Broutin, Neininger and Sulzbach used the "contraction method" to obtain a convergence in distribution as $t \to \infty$ of the rescaled processes $\{t^{-\beta} N_t(x) : 0 \le x \le 1\}$ towards a random continuous process $\{\tilde{M}_\infty(x) : 0 \le x \le 1\}$ characterized by a recursive decomposition. The main result of the present work is to show that this convergence actually holds in a stronger sense:

Theorem 1. *For every $x \in [0,1]$ we have the following* almost sure *convergence*

$$t^{-\beta} N_t(x) \quad \xrightarrow[t\to\infty]{a.s.} \quad K_0 \cdot \tilde{M}_\infty(x).$$

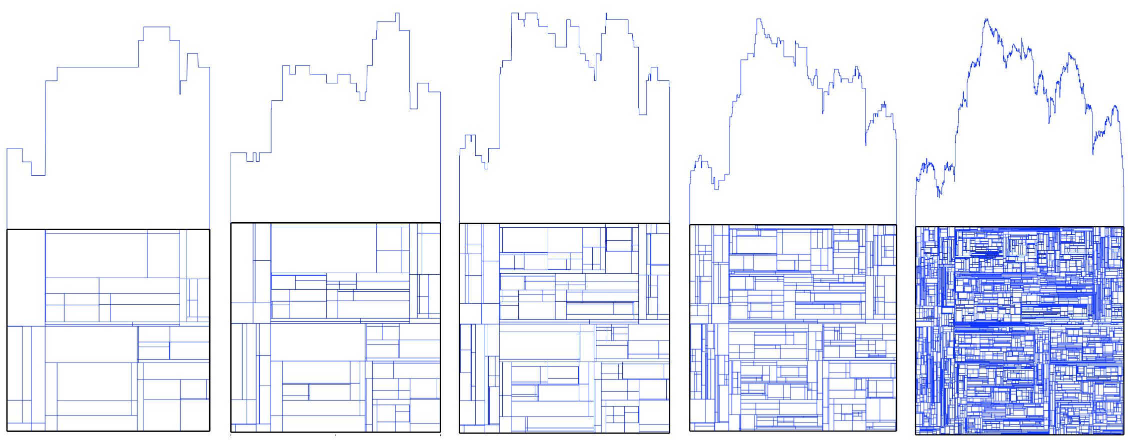

Figure 2: An illustration of the strong convergence of the partial match queries. The curves above the quadtree represent the renormalized processes $t^{-\beta}(N_t(x))_{x\in[0,1]}$ for $t = 20, 50, 100, 500$ and 3000.

The random variable $\tilde{M}_\infty(x)$ is obtained as the limiting value of a continuous-time martingale $\{M_t(x) : t \ge 0\}$ defined by (2) which is a variation on the martingale introduced in [4, Section 3.2]. The martingale $M_t(x)$ must be considered as an analogous of the well-known Malthusian martingale in fragmentation theory, see [1]. Indeed, the convergence of $t^{-\beta} N_t(x)$ towards the limiting value of $M_t(x)$ is similar to the proof of [2, Theorem 5] and requires some of the estimates of [4]. See Section 5 for comments.

Here is a direct corollary of Theorem 1 combined with the results of [4]:

Corollary 2. *We have the following convergence in probability*

$$\left(t^{-\beta} N_t(x)\right)_{x\in[0,1]} \quad \xrightarrow[t\to\infty]{(P)} \quad K_0 \cdot \left(\tilde{M}_\infty(x)\right)_{x\in[0,1]},$$

for the uniform metric $\|.\|_\infty$.

The note is organized as follows: We first introduce the martingales whose limit value furnishes the process $\{\tilde{M}_\infty(x) : 0 \le x \le 1\}$ and recall some of its properties. The third section is devoted to an estimate on the smallest and the largest rectangle in the quadtree at time $t > 0$ which is used in the proof of the main result. In the last section we give some comments related to fragmentation theory.

Acknowledgment: I am grateful to Nicolas Broutin, Ralph Neininger and Henning Sulzbach for keeping me informed about their recent work on quadtrees. Special thanks go to Adrien Joseph and Henning Sulzbach for precious comments on a first version of this work.

2 The martingales

In this section we introduce the martingale which the proof of Theorem 1 is based on and compare it with the one introduced in [4, Section 3.2]. We start by setting some notation.

Recall the definitions of β and of the map h given in the Introduction. The genealogy of the rectangles appearing in the quadtree process $(\mathrm{Quad}(t))_{t\ge 0}$ can be encoded on the full infinite 4-ary tree

$$\mathcal{T}_4 \quad := \quad \bigcup_{n\ge 0} \{1,2,3,4\}^n.$$

The first square $[0,1]^2$ corresponds to the word $\varnothing \in \mathcal{T}_4$ and when a rectangle encoded by a word $u \in \mathcal{T}_4$ is split, we encode the four resulting subrectangles by $u1, u2, u3$ and $u4$ in counter clockwise order starting with the north-east rectangle. This genealogy induces a notion of ancestor, offspring... on the rectangles of $\cup\mathrm{Quad}(t)$. The generation of a rectangle R that appears in the quadtree process is the length of its encoding word in $\mathcal{T}_4$ and is denoted by $\mathrm{Gen}(R)$ (the length of $\varnothing$ is 0 by convention). The time of appearance of R is the first $t > 0$ such that $R \in \mathrm{Quad}(t)$ and is denoted by $\mathrm{Time}(R)$.

For $t > 0$ and $x \in [0,1]$ we denote by $\mathcal{Q}_t(x) := \{Q_t^i(x)\}_{i\ge 1}$ the set of all rectangles belonging to $\mathrm{Quad}(t)$ whose first coordinate intersects x. The left-most and right-most horizontal coordinates of $Q_t^i(x)$ are denoted by ℓ_t^i and r_t^i and we write

$$x_t^i \quad := \quad \frac{x - \ell_t^i}{r_t^i - \ell_t^i},$$

for the "position" of x inside the rectangle $Q_t^i(x)$. By standard properties of the Poisson point process Π, conditionally on the sigma-field $\mathcal{F}_t$ generated by $(\mathrm{Quad}(u))_{0\le u\le t}$ the number of rectangles in $\mathcal{Q}_{t+s}(x)$ whose ancestor at time t is $Q_t^i(x)$ has the same distribution as

$$N'_{s\cdot\lambda(Q_t^i(x))}(x_t^i),$$

where λ stands for the two-dimensional Lebesgue measure on $[0,1]^2$ and $N'_{\cdot}(.)$ is an independent copy of the process $N_{\cdot}(.)$.

Proposition 3. *For every* $x \in [0,1]$, *the process*

$$t \quad \longmapsto \quad M_t(x) \quad := \quad \sum_{i\ge 1} \lambda\left(Q_t^i(x)\right)^\beta h\left(x_t^i\right), \tag{2}$$

is a continuous-time non-negative martingale whose limiting value is denoted by $\tilde{M}_\infty(x)$*.*

Before going into the proof of Proposition 3 let us emphasize the difference between this martingale and the one considered in [4]. Fix a generation $n \geq 0$ and denote by $\{\tilde{Q}^i_n(x)\}_{1 \leq i \leq 2^n}$ the rectangles at generation n that are above the point $x \in [0,1]$ and write $\tilde{x}^i_n$ for the position of x inside $\tilde{Q}^i_n(x)$. Then from [4, Section 3.2],

$$\tilde{M}_n(x) \quad := \quad \sum_{i=1}^{2^n} \lambda\big(\tilde{Q}^i_n(x)\big)^\beta h(\tilde{x}^i_n),$$

is a discrete-time non-negative martingale whose limiting value is $\tilde{M}_\infty(x)$. The proof of this fact is based on the following lemma that shows that the expectation is kept after one splitting.

Lemma 4. *We have*

$$E\left[\sum_{i=1}^{2} \lambda(\tilde{Q}^i_1(x))^\beta h(\tilde{x}^i_1)\right] \quad = \quad h(x).$$

This lemma was proved in [4] but is also rigorously equivalent to the fact h solves the integral equation that was already considered in [5, Section 5].

The difference between the martingales $\tilde{M}_n(x)$ and $M_t(x)$ is that in the latter case we consider the splittings chronologically as they occur whereas in the first case we consider them generation after generation. It should be clear that the order in which the splittings are considered does not change the martingale property and we could use Lemma 4 to show directly that $M_t(x)$ is a martingale for every $x \in [0,1]$. However it will be useful for our purpose to link $M_t(x)$ to its discrete time analog $\tilde{M}_n(x)$.

Proof of Proposition 3. Fix $x \in [0,1]$. It easily follows from [4, Section 3.2] (see also [3]) that $\tilde{M}_n(x)$ converges in L^p for any $p > 1$ towards $\tilde{M}_\infty(x)$ and thus $E[\tilde{M}_\infty(x) \mid \mathcal{F}_t] = \lim E[\tilde{M}_n(x) \mid \mathcal{F}_t]$ almost surely as $n \to \infty$. By the Markov property applied at time $t > 0$ and using the martingale structure of $\tilde{M}_n(x)$ we deduce that

$$E[\tilde{M}_n(x) \mid \mathcal{F}_t] \quad = \quad \sum_{i \geq 1} \lambda(Q^i_t(x))^\beta h(x^i_t) \mathbf{1}_{\mathrm{Gen}(Q^i_t(x)) < n} + \sum_{i=1}^{2^n} \lambda(\tilde{Q}^i_n(x))^\beta h(\tilde{x}^i_n) \mathbf{1}_{\mathrm{Time}(\tilde{Q}^i_n(x)) \leq t}.$$

It is easy to see that $\inf\{\mathrm{Time}(\tilde{Q}^i_n(x)) : 1 \leq i \leq 2^n\}$ goes to infinity as $n \to \infty$ a.s., hence letting n tend to infinity in the last display we get that $E[\tilde{M}_\infty(x) \mid \mathcal{F}_t] = M_t(x)$ and $M_t(x)$ is a non-negative continuous-time martingale that converges almost surely and in any L^p for $p > 1$ towards $\tilde{M}_\infty(x)$. This completes the proof of the proposition. $\square$

The process $x \in [0,1] \mapsto \tilde{M}_\infty(x)$ was used in [4] to construct a fixed point to a recursive equation in distribution. In particular it is proved that $x \mapsto \tilde{M}_\infty(x)$ is almost surely continuous.

3 A geometric estimate

In this section we establish a rough control on the area of the largest and the smallest rectangle of $\mathrm{Quad}(t)$. The reader can skip this part on first reading. For $t > 0$, let $I_t := \inf \lambda(R)$ and $S_t := \sup \lambda(R)$ where the infimum and supremum run over all the rectangles $R \in \mathrm{Quad}(t)$. We will roughly prove that $t^{-4+o(1)} \leq I_t$ and $S_t \leq t^{-1+o(1)}$. The formal statement is the following:

Lemma 5. *For every $\varepsilon > 0$ we have*

$$-4-\varepsilon < \liminf_{t\to\infty} \frac{\log\left(I_t\right)}{\log(t)} \quad a.s. \qquad and \qquad \limsup_{t\to\infty} \frac{\log\left(P(S_t > t^{-1+\varepsilon})\right)}{\log(t)} = -\infty.$$

Proof. Lower bound. Let $(x_i)_{1\le i\le n}$ and $(y_i)_{1\le i\le n}$ be the coordinates of the points of Π that occur before time t. By standard properties of Poisson point processes, conditionally on n, (x_i) and (y_i) are independent sequences of n i.i.d. uniform variables over $[0,1]$. A simple geometric argument (see Fig. 3 below) shows that

$$I_t \quad \ge \quad \min_{\substack{i\neq j\\ 1\le i,j\le n}} |x_i - x_j| \cdot \min_{\substack{i\neq j\\ 1\le i,j\le n}} |y_i - y_j|.$$

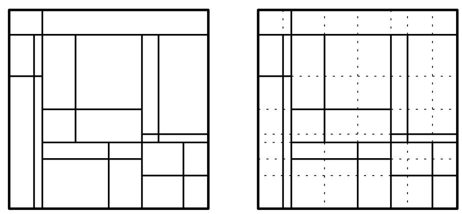

Figure 3: Illustration of the lower bound.

Let $\varepsilon > 0$. By classical results on the uniform sieve of the interval $[0,1]$, if $x_1, x_2, ...$ are i.i.d. uniform points over $[0,1]$ then $\min\{|x_i - x_j| : 1 \le i,j \le n, i \neq j\}$ is asymptotically larger than $n^{-2-\varepsilon}$ a.s.. Indeed we have

$$P\left(x_n \in \bigcup_{i=1}^{n-1} [x_i - n^{-2-\varepsilon}, x_i + n^{-2-\varepsilon}]\right) \quad \le \quad 2n^{-1-\varepsilon},$$

and an application of Borel Cantelli's lemma proves the claim. Since eventually $t/2 \le n \le 2t$ and $t \mapsto I_t$ is decreasing we almost surely have $I_t \ge t^{-4-2\varepsilon}$ eventually.

Upper bound. We use a common technique in fragmentation theory: the tagged particle. Assume that independently of the quadtree process $(\mathrm{Quad}(t))_{t\ge 0}$ we are given an independent variable (U,V) uniformly distributed over $[0,1]^2$. The rectangle $R_t^\bullet \in \mathrm{Quad}(t)$ containing (U,V) is called the "tagged rectangle" at time t. The distribution of $(R_t^\bullet)_{t\ge 0}$ is equivalently described as follows. We start with $R_0^\bullet := [0,1]^2$ and define the process $R_t^\bullet$ iteratively: when a splitting occurs at time τ inside the tagged rectangle $R_{\tau-}^\bullet$, then $R_\tau^\bullet$ is one of the four subrectangles of $R_{\tau-}^\bullet$ chosen proportionally to its two-dimensional Lebesgue measure.

It is clear from the above construction that the tagged rectangle at generation n has a two-dimensional Lebesgue measure which is distributed according to

$$\prod_{i=1}^{2n} \overline{U}_i,$$

where $\overline{U}_i$ are independent identically distributed variables with density $2m\mathbf{1}_{0<m<1}\mathrm{d}m$ (size-biased uniform over $[0,1]$). In particular we have $P(\prod_{i=1}^{2n} \overline{U}_i \ge \alpha^n) \le E[\overline{U}_i]^{2n}\alpha^{-n} = (4/9\alpha)^n$. We now turn to the study of the mass of the tagged rectangle at time $t > 0$. For every $\varepsilon > 0$ we have

$$P(\lambda(R_t^\bullet) > t^{-1+2\varepsilon}) \quad \le \quad P(\mathrm{Gen}(R_t^\bullet) \le t^\varepsilon\,,\ \lambda(R_t^\bullet) > t^{-1+2\varepsilon}) + P(\mathrm{Gen}(R_t^\bullet) > t^\varepsilon\,,\ \lambda(R_t^\bullet) > t^{-1+2\varepsilon}).$$

By our preceding remark, for large $t > 0$, the second term of the right-hand side is bounded above by $2^{-\lfloor t^\varepsilon \rfloor}$. For the first term, remark that if $\lambda(R_t^\bullet) > t^{-1+2\varepsilon}$ then for every $0 \le s \le t$, the intensity at which a particle falls inside $R_s^\bullet$ is larger than $t^{-1+2\varepsilon}$, thus by standard properties of exponential variables we have $P(\mathrm{Gen}(R_t^\bullet) \le t^\varepsilon\,,\ \lambda(R_t^\bullet) > t^{-1+2\varepsilon}) \le P(\mathcal{P}(t^{2\varepsilon}) \le t^\varepsilon)$ where $\mathcal{P}(t^{2\varepsilon})$ is a Poisson distribution of mean $t^{2\varepsilon}$. Let us make this more precise. For $n \ge 0$, denote $\tilde{R}_n^\bullet$ the tagged rectangle at generation n. The rectangle $\tilde{R}_n^\bullet$ thus lives for an exponential time of parameter $\lambda(\tilde{R}_n^\bullet)$ before it splits. We deduce that if $\mathcal{E}_0, \mathcal{E}_1, \ldots$ is an i.i.d. sequence of exponential variables of parameter one which is also independent of $\lambda(\tilde{R}_0^\bullet), \lambda(\tilde{R}_1^\bullet), \ldots$ then

$$\begin{aligned}
P(\mathrm{Gen}(R_t^\bullet) \le t^\varepsilon\,,\ \lambda(R_t^\bullet) > t^{-1+2\varepsilon}) &= P\left(\sum_{i=0}^{\mathrm{Gen}(R_t^\bullet)} \lambda(\tilde{R}_i^\bullet)^{-1} \cdot \mathcal{E}_i > t\,,\ \lambda(\tilde{R}_t^\bullet) > t^{-1+2\varepsilon}\,,\ \mathrm{Gen}(R_t^\bullet) \le t^\varepsilon\right) \\
&\le P\left(t^{1-2\varepsilon} \sum_{i=0}^{\mathrm{Gen}(R_t^\bullet)} \mathcal{E}_i > t\,,\ \mathrm{Gen}(R_t^\bullet) \le t^\varepsilon\right) \\
&\le P\left(\sum_{i=0}^{\lfloor t^\varepsilon \rfloor} \mathcal{E}_i > t^{2\varepsilon}\right) \\
&= P\big(\mathcal{P}(t^{2\varepsilon}) \le t^\varepsilon\big).
\end{aligned}$$

The last probability being bounded above by $c^{-1}\exp(-ct^\varepsilon)$ for some $c > 0$. Gathering-up the pieces, there exists $c' > 0$ such that we have $P(\lambda(R_t^\bullet) > t^{-1+2\varepsilon}) \le c'^{-1}\exp(-c't^\varepsilon)$ for every $t > 0$. We then use the tagged fragment to bound S_t from above. Notice that at any time $t > 0$ the tagged fragment $R_t^\bullet$ has a probability S_t of being the largest fragment, thus

$$P(\lambda(R_t^\bullet) > t^{-1+2\varepsilon}) \quad \ge \quad P(S_t > t^{-1+2\varepsilon})t^{-1+2\varepsilon},$$

which together with the previous bound easily completes the proof of the lemma.

□

4 Proof of the main results

4.1 Proof of Theorem 1

Proof. Let us first describe the main idea of the proof, which is similar to [2, Theorem 5] and roughly speaking reduces to apply a law of large number after conditioning at a large time $t > 0$. Fix $x \in [0,1]$ and let T be much larger than t. Conditionally on $\mathcal{F}_t$ the variable $N_T(x)$ is the sum of $N_t(x)$ independent contributions corresponding to the offsprings of the rectangles above x at time t. By standard properties of the quadtree construction, the number of descendants of the rectangle $Q_t^i(x)$ inside $\mathcal{Q}_T^i(x)$ is distributed as $N'_{\lambda(Q_t^i(x))(T-t)}(x_t^i)$ where $N'_\cdot(.)$ is an independent copy of the process $N_\cdot(.)$. Thus if for $x \in [0,1]$ and $t \ge 0$ we set $f(t,x) := E[N_t(x)]$, we have

$$E\big[N_T(x) \mid \mathcal{F}_t\big] \quad = \quad \sum_{i \ge 1} f\big(\lambda(Q_t^i(x))(T-t), x_t^i\big).$$

We now let $T \to \infty$ in the last display. Since for each rectangle $Q_t^i(x)$ of $\mathcal{Q}_t(x)$ we have $(T-t)\lambda(Q_t^i(x)) \to \infty$ then $T^{-\beta} f(\lambda(Q_t^i(x))(T-t), x_t^i)$ tends to $K_0 \lambda(Q_t^i)^\beta h(x_t^i)$ as $T \to \infty$.

Henceforth we have

$$\begin{aligned}
&\left|T^{-\beta}E\big[N_T(x)\mid\mathcal{F}_t\big]-K_0\cdot\sum_{i\geq 1}\lambda(Q_t^i(x))^{\beta}h(x_t^i)\right|\\
=\;&\left|T^{-\beta}E\big[N_T(x)\mid\mathcal{F}_t\big]-K_0\cdot M_t(x)\right| \quad \xrightarrow[T\to\infty]{} \quad 0. \qquad (3)
\end{aligned}$$

The strategy of the proof is now clear: conditionally on $\mathcal{F}_t$, by the law of large numbers, $T^{-\beta}N_T(x)$ will be very close to its (conditional) mean which is close to $K_0\cdot M_t(x)$ which converges towards $K_0\cdot\tilde{M}_\infty$. This will imply the theorem.

To make this statement precise, and in particular get an almost sure convergence (a convergence in probability would be much easier to prove), we shall need the estimates on the expectation and the variance of the process $N_t(x)$ proved by Broutin, Neininger and Sulzbach. It follows from Proposition 11 and Theorem 5 in [4] that there exist two constants $C>0$ and $\delta>0$ such that for every $t>0$ we have

$$\sup_{0\leq x\leq 1}\left|t^{-\beta}E[N_t(x)]-K_0\cdot h(x)\right| \;\leq\; Ct^{-\delta} \qquad (4)$$

$$\sup_{0\leq x\leq 1}\big(\mathrm{Var}(N_t(x))\big) \;\leq\; C(t^{2\beta}+t), \qquad (5)$$

the term t appearing in the last line since the variance of $N_t(x)$ is of order t near $t=0$. We first make (3) quantitative in T. Fix $\alpha\geq 5$ such that $\delta(\alpha-5)>1$ and for $t>0$, set $T:=t^\alpha$. By the choice of t versus T and the lower bound of Lemma 5 we get that almost surely, there exists a random time τ such that for $t\geq\tau$ we have $(T-t)\inf\{\lambda(Q_t^i(x)) : i\geq 1\}\geq t^{\alpha-5}$. Henceforth using the bound (4), we have for $t\geq\tau$

$$\begin{aligned}
\left|(T-t)^{-\beta}E\big[N_T(x)\mid\mathcal{F}_t\big]-K_0\cdot M_t(x)\right| \;&\leq\; \sum_{i\geq 1}\left|(T-t)^{-\beta}f\big(\lambda(Q_t^i(x))(T-t),x_t^i\big)-K_0\cdot\lambda(Q_t^i(x))^{\beta}h(x_t^i)\right|\\
&\leq\; CN_t(x)t^{-\delta(\alpha-5)}.
\end{aligned}$$

Since $N_t(x)$ is clearly less that the number of points fallen so far, the definition of α implies that $N_t(x)t^{-\delta(\alpha-5)}$ goes to 0 almost surely. Since $M_t(x)$ is almost surely bounded, we proved that with our choice of α we have $|T^{-\beta}E[N_T(x)\mid\mathcal{F}_t]-K_0\cdot M_t(x)|\to 0$ almost surely as $t\to\infty$ and using Proposition 3 if follows that

$$\left|T^{-\beta}E[N_T(x)\mid\mathcal{F}_t]-K_0\cdot\tilde{M}_\infty(x)\right| \quad \xrightarrow[t\to\infty]{a.s.} \quad 0. \qquad (6)$$

Recall that conditionally on $\mathcal{F}_t$ the contributions of each rectangle $Q_t^i(x)$ to $N_T(x)$ are independent, thus we have

$$\begin{aligned}
T^{-2\beta}E\left[\big(N_T(x)-E[N_T(x)\mid F_t]\big)^2\mid\mathcal{F}_t\right] \;&=\; T^{-2\beta}\sum_{i\geq 1}\mathrm{Var}\big(N_{\lambda(Q_t^i(x))(T-t)}(x_t^i)\big)\\
&\leq\; C\Big(\sum_{i\geq 1}\lambda(Q_t^i(x))^{2\beta}+T^{1-2\beta}\sum_{i\geq 1}\lambda(Q_t^i(x))\Big)\\
&\leq\; C\big(S_t^{2\beta-1}+T^{1-2\beta}\big)\sum_{R\in\mathrm{Quad}(t)}\lambda(R)\\
&=\; C(S_t^{2\beta-1}+T^{1-2\beta}),
\end{aligned}$$

where we used the bound (5) to reach the second line and the fact that $2\beta>1$ to go from the third to the last line. Let $\varepsilon>0$. Applying the standard Markov inequality conditionally on $\mathcal{F}_t$ we obtain

$$P\Big(\,T^{-\beta}\big|N_T(x)-E[N_T(x)\mid\mathcal{F}_t]\big|\geq\varepsilon\;\big|\;\mathcal{F}_t\Big) \;\leq\; C\varepsilon^{-2}\big(S_t^{2\beta-1}+T^{1-2\beta}\big). \qquad (7)$$

We now take $T_k := (1+\eta)^k$ and $t_k := (1+\eta)^{k/\alpha}$ for $k = 1,2,3...$ and $\eta > 0$. Since $E[S_{t_k}^{2\beta-1}]$ is less than $t_k^{(2\beta-1)(\varepsilon-1)} + P(S_{t_k} > t_k^{-1+\varepsilon})$, using Lemma 5 and (7) we see that $P(T_k^{-\beta}|N_{T_k}(x) - E[N_{T_k}(x) \mid \mathcal{F}_{t_k}]| \geq \varepsilon)$ is summable in k. Applying Borel-Cantelli's lemma and using (6) we deduce that for every $\eta > 0$ we have the following almost sure convergence

$$(1+\eta)^{-k\beta} N_{(1+\eta)^k}(x) \quad \xrightarrow[k\to\infty]{a.s.} \quad K_0 \cdot \tilde{M}_\infty(x).$$

To extend this result to the whole process we use the fact that $t \mapsto N_t(x)$ is increasing in t which implies

$$(1+\eta)^{-(k+1)\beta} N_{(1+\eta)^k}(x) \leq s^{-\beta} N_s(x) \leq (1+\eta)^{-k\beta} N_{(1+\eta)^{k+1}}(x)$$

for every $(1+\eta)^k \leq s \leq (1+\eta)^{k+1}$ and $k \geq 1$. Since this holds for any $\eta > 0$ we easily deduce that $t^{-\beta} N_t(x)$ almost surely converges towards $K_0 \cdot \tilde{M}_\infty(x)$. This completes the proof of Theorem 1. □

4.2 Proof of Corollary 2

Proof. (Sketch) Theorem 1 implies the convergence of the finite dimensional marginals of $t^{-\beta} N_t(.)$ towards those of $K_0 \cdot \tilde{M}_\infty(.)$ in probability: For any $0 \leq x_1, ..., x_k \leq 1$ we have

$$t^{-\beta}\big(N_t(x_i)\big)_{1\leq i\leq k} \quad \xrightarrow[t\to\infty]{(P)} \quad K_0\big(\tilde{M}_\infty(x_i)\big)_{1\leq i\leq k}. \tag{8}$$

Furthermore Theorem 1 of [4] provides the tightness of the processes $t^{-\beta}(N_t(.))$ for the uniform metric: for every $\varepsilon > 0$ there exists $\eta > 0$ such that for $t > 0$ large enough we have

$$P(\omega_{t^{-\beta}N_t(.)}(\eta) \leq \varepsilon) \geq 1 - \varepsilon, \tag{9}$$

where $\omega_g(\eta) = \sup\{|g(x) - g(y)|\,,\ |x-y| \leq \eta\}$ is the modulus of continuity of the function g. Recalling that $x \mapsto \tilde{M}_\infty(x)$ is almost surely continuous, we can combine (8) and (9) to get that $t^{-\beta} N_t(.)$ converges in probability for the L^∞ metric towards $\tilde{M}_\infty(.)$. We leave the details to the reader. □

Open Question. It is believable that the convergence of Corollary 2 actually holds almost surely, that is

$$\big(t^{-\beta} N_t(x)\big)_{x\in[0,1]} \quad \xrightarrow[t\to\infty]{a.s.} \quad K_0 \cdot \big(\tilde{M}_\infty(x)\big)_{x\in[0,1]},$$

for the uniform metric $\|.\|_\infty$.

5 Fragmentation process with parameter

In this section we comment at an informal level on the strategy adopted in this work and on possible extensions of our techniques.

Fragmentation theory. Let us briefly recall some basics about fragmentation theory. We stick to a very simple case for sake of clarity. For more details, we refer to [1]. To define a self-similar fragmentation process[1] $\mathscr{F}$ we need one input: a probability measure ν on $\{(s_1, s_2) : s_1 \geq s_2 > 0 \text{ and } s_1 + s_2 \leq 1\}$. The process $\mathscr{F}$ with values in the set $\mathcal{S}^\downarrow = \{(s_1, s_2, \dots) : s_1 \geq s_2 \geq \cdots \geq 0 \text{ and } \sum_i s_i \leq 1\}$ is informally characterized as follows:

[1]binary, without erosion and with dislocation measure of mass one

if at time t we have $\mathscr{F}(t) = (s_1(t), s_2(t), \dots)$, then for every $i \geq 1$, the i-th "particle" of mass $s_i(t)$ lives an exponential time with parameter $s_i(t)$ before splitting into two particles of masses $r_1 s_i(t)$ and $r_2 s_i(t)$, where (r_1, r_2) has been sampled from ν independently of the past and of the other particles. In other words, each particle undergoes a self-similar fragmentation with time rescaled by its mass. It is classical that under mild assumption there exists a unique $\beta \in (0,1]$ (called the Malthusian exponent) such that

$$\int \mathrm{d}\nu(s_1, s_2)\ s_1^\beta + s_2^\beta \quad = \quad 1, \tag{10}$$

and that $M_t := \sum s_i(t)^\beta$ is a continuous-time non-negative martingale which plays a central role in the asymptotic behavior of the fragmentation process, see [1, 2].

Parametrized fragmentation. In the problem of the partial match query, one can think of the rectangles above the point x at time $t > 0$ as a *fragmentation process where the particles have an additional parameter*, in our case the position $x_t^i \in [0,1]$. This leads us to extend the notion of dislocation measure and to define a fragmentation process with parameter: a parametrized (binary) dislocation probability is a collection $\boldsymbol{\nu} = (\nu_x)_{x\in[0,1]}$ such that for every $x \in [0,1]$, ν_x is a probability measure on

$$\{(s_1, x_2, s_2, x_2) \in [0,1]^4 : s_1 \geq s_1 \text{ and } s_1 + s_2 \leq 1\}.$$

A parametrized fragmentation process $\mathcal{F}$ with dislocation measure $\boldsymbol{\nu}$ is then a process with values in $\{(s_1, x_1, s_2, x_2, \dots\dots) \in [0,1]^{\mathbb{N}} : s_1 \geq s_2 \geq \dots \geq 0 \text{ and } \sum_i s_i \leq 1\}$ whose evolution is informally described as follows: We start with a particle of mass 1 given with a position $x \in [0,1]$. If $\mathcal{F}(t) = (s_1(t), x_1(t), s_2(t), x_2(t), \dots)$ then for every $i \geq 1$, the i-th "particle" of mass $s_i(t)$ with position $x_i(t)$ lives an exponential time with parameter $s_i(t)$ before splitting into two particles of masses $r_1 s_i(t)$ and $r_2 s_i(t)$ with respective positions y_1 and y_2, where (r_1, y_1, r_2, y_2) has been sampled from $\nu_{x_i(t)}$ independently of the past and of the other particles. In this setting, (10) is replaced by the following assumption: (H) There exists $\beta \in [0,1]$ and $h : x \in [0,1] \mapsto h(x) \in \mathbb{R}_+$ such that for every $x \in [0,1]$ we have

$$\int \mathrm{d}\nu_x(s_1, x_1, s_2, x_2)\big(s_1^\beta h(x_1) + s_2^\beta h(x_2)\big) \quad = \quad h(x). \tag{11}$$

Then under this assumption the process $M_t(x) := \sum s_i(t)^\beta h(x_i(t))$ is a continuous-time non-negative martingale playing the role of the Malthusian martingale. It is believable that substantial parts of self-similar fragmentations theory can be adapted to this parametrized case.

Examples. PARTIAL MATCH QUERIES IN QUADTREE. Within this formalism the process of the masses of the rectangles of $\mathcal{Q}_t(x)$ is a parametrized fragmentation process starting with a single particle of mass 1 and parameter x. Its parametrized dislocation measure $\boldsymbol{\nu}^{\mathbf{quad}}$ is given by

$$\begin{aligned}
&\int \mathrm{d}\nu_x^{\text{quad}}(s_1, x_1, s_2, x_2) f(s_1, x_1, s_2, x_2)\\
= \quad &\iint_{[0,1]^2} \mathrm{d}u\mathrm{d}v \left(\mathbf{1}_{x<u} f\left(uv, \frac{x}{u}, u(1-v), \frac{x}{u}\right) + \mathbf{1}_{x>u} f\left((1-u)v, \frac{x-u}{1-u}, (1-u)(1-v), \frac{x-u}{1-u}\right)\right),
\end{aligned}$$

for every $x \in [0,1]$ and every Borel function $f : [0,1]^4 \to \mathbb{R}_+$. In particular $\mathrm{d}\nu_{\text{quad}}(s_1, x_1, s_2, x_2)$-almost surely we have $x_1 = x_2$ and Hypothesis (H) is fulfilled with $\beta = \frac{\sqrt{17}-3}{2}$ and $h(x) = (x(1-x))^{\beta/2}$.

To conclude, besides the application of the method to the problem of partial match queries in higher-dimensional random quadtrees or in random k-d trees, we present another setup taken from [6] where the concept of "parametrized fragmentation" could be applied (although the results there only rely on the standard fragmentation theory).

Random chords. We recall the random chord construction of [6]. We consider a sequence $U_1, V_1, U_2, V_2, \ldots$ of independent random variables, which are uniformly distributed over the unit circle $\mathbb{S}_1$. We then construct inductively a sequence $L_1, L_2, \ldots$ of random closed subsets of the (closed) unit disk $\overline{\mathbb{D}}$. To begin with, L_1 just consists of the chord with endpoints U_1, and V_1, which we denote by $[U_1 V_1]$. Then at step $n+1$, we consider two cases. Either the chord $[U_{n+1}V_{n+1}]$ intersects L_n, and we put $L_{n+1} := L_n$. Or the chord $[U_{n+1}V_{n+1}]$ does not intersect L_n, and we put $L_{n+1} := L_n \cup [U_{n+1}V_{n+1}]$. Thus, for every integer $n \geq 1$, L_n is a disjoint union of random chords. If $x, y \in \mathbb{S}_1$ then one defines the fragments separating x from y as the connected components of $\overline{\mathbb{D}} \backslash L_n$ intersecting $[x, y]$.

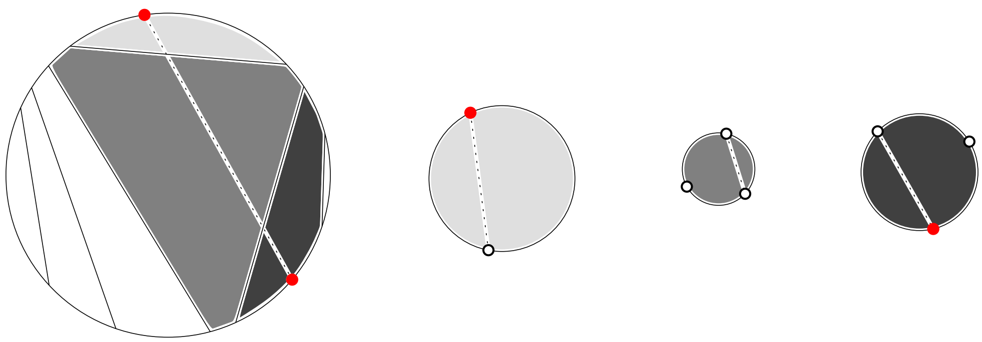

Figure 4: The fragments separating two points.

After contracting the chords of L_n, each fragment F separating x from y can be seen as a particle with two distinguished points (see Fig. 4) whose mass corresponds to the one-dimensional Lebesgue measure of $F \cap \mathbb{S}_1$, see [6]. The "position" or parameter of each particle is then the relative positions of its two distinguished points in $[0, 1]$. It was shown in [6] that if U, V are independent and uniformly distributed over $\mathbb{S}_1$ then the fragments separating U from V form (in a proper continuous time parametrization) a fragmentation process. In the case when $x, y \in \mathbb{S}_1$ are fixed, the process of fragments separating x from y (in a proper continuous time parameterization) can be seen as a parametrized fragmentation process with assumption (H) fulfilled with

$$\beta = \frac{\sqrt{17} - 3}{2} \quad \text{and} \quad h(x) = (x(1-x))^{\beta},$$

which is equivalent to equation (14) of [6].

References

[1] J. Bertoin. *Random Fragmentations and Coagulation Processes.* Number 102 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2006.

[2] J. Bertoin and A. Gnedin. Asymptotic laws for nonconservative self-similar fragmentations. *Electron. J. Probab.*, 9(19):575–593, 2004.

[3] N. Broutin, R. Neininger, and H. Sulzbach. Asymptotic analysis of Partial Match retrieval *In preparation*, 2011.

[4] N. Broutin, R. Neininger, and H. Sulzbach. Partial match queries in random quadtrees. *arXiv:1107.2231*, 2011.

[5] N. Curien and A. Joseph. Partial match queries in random quadtrees : A probabilistic approach. *Adv. in Appl. Probab.*, 43:178–194, 2011.

[6] N. Curien and J.-F. Le Gall. Random recursive triangulations of the disk *via* fragmentation theory. *Ann. Probab.* to appear.

[7] R. A. Finkel and J. L. Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, mars 1974.

[8] P. Flajolet, G. Gonnet, C. Puech, and J. M. Robson. Analytic variations on quadtrees. *Algorithmica*, 10(6):473–500, 1993.

[9] P. Flajolet and R. Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009.

Département de Mathématiques et Applications
École Normale Supérieure, 45 rue d'Ulm
75230 Paris cedex 05, France

nicolas.curien@ens.fr